RNA: nuclear glue for folding the genome

RNA: nuclear glue for folding the genome

Ryu-Suke Nozawa and Nick Gilbert

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Rd

Edinburgh, EH4 2XU, UK

Correspondence: nick.gilbert@ed.ac.uk (N.Gilbert)

Keywords

hnRNA, Chromatin, Nuclear Architecture, Chromatin-associated RNAs, RNA debris, Transcription

Abstract

A significant amount of RNA is present in the nucleus of mammalian cells but only a small proportion of it is destined for the cytoplasm and subsequent translation, leaving much RNA to associate with chromatin. Historically nuclear RNA was thought to interact with proteins to form a filamentous nuclear matrix, but this idea became less popular as more dynamic models of chromatin behavior became more prevalent. Using new molecular and imaging approaches it is becoming clear that RNA should be considered as an integral component of nuclear organisation; it is transcriptionally responsive and interacts with abundant nuclear RNA binding proteins. We suggest that these protein/RNA structures form a dynamic nuclear mesh that can regulate interphase chromatin structure.

Nuclear RNAs

In mammalian cells DNA is protected from damage by packaging with histone proteins to form chromatin. This maintains genetic integrity and enables our blueprint to be passed to daughter cells intact. Live cell imaging shows mammalian chromatin is mobile in early G1 but once chromosome territories have adopted their preferred positions they are relatively static and instead undergo constrained diffusion. In contrast to DNA, the other nucleic acid found in the nucleus, RNA, is dynamic and readily turned over. Unlike DNA there are also many different types of RNA which have multiple roles throughout the nucleus and into the cytoplasm (see Table 1). Nuclear RNAs are highly abundant and consist of species with defined function whilst others are often thought of as by-products of other nuclear processes, or as RNA debris. In the past a stable **nuclear matrix** (see Glossary) [1], rich in RNA and protein, underpinning nuclear organization was hypothesized, but the idea was abandoned as more dynamic models of chromatin behaviour became prevalent. In this review we discuss recent data and ideas that suggest nuclear RNAs interact with proteins [2,3] to form a dynamic nuclear mesh important for folding the genome, updating our ideas on nuclear organisation.

hnRNAs

Heterogenous nuclear RNAs (**hnRNAs**) are a diverse class of different RNA species that encompass both coding, non-coding and regulatory RNAs (Table 1). Going back 30 years cellular RNA was typically monitored using radio-labelling (Box 1), pulse labelling was used to track both its synthesis, composition, and where in the nucleus it was transcribed. Seminal studies showed transcription was concentrated on the surface of protein-rich “factories” [4-6], which we now know from super-resolution imaging studies are highly dynamic [7] and are rich in nascent and non-coding RNA [8]. It is often forgotten that only 5% of hnRNA reaches the cytoplasm, many of these will be polyadenylated and destined for making proteins, whilst the remainder of RNAs presumably remain inside the nucleus either performing a function or being degraded by ribonucleases [9]. Although early experiments provided crucial insights into the rate RNAs were being produced and their distribution [10] they don’t give an impression of the sequence or function of different RNAs.

Surprisingly, quantitative experiments in the 1960s and ‘70s suggested that as much as 10% of chromatin by mass was RNA [11]. To visualise these structures protein and DNA were extracted leaving an RNA-rich filamentous structure, visible by electron microscopy, and termed a nuclear matrix, which was reported to consist of over 75% hnRNA [12,13]. However, there were many disagreements within the field as to whether these structures were real or instead were an artefact of the preparation. Experimental conditions were often harsh, and it was often joked there were more methods for preparing a nuclear matrix than there were publications! Towards the end of the nuclear matrix heyday two key pieces of

experimental evidence seemed to be incompatible with the idea of a stable nuclear framework. Firstly, many proteins were fluorescently tagged with GFP and none seemed to show a matrix-like organisation. Secondly, the mobility of many of these GFP tagged proteins were analysed by **FRAP** and were found to be highly dynamic; an idea that seemed inconsistent with a stable nuclear matrix. However, many of the images of extracted nuclei were compelling and strongly supported the notion that RNA was a key player in nuclear organisation. Still, it was not clear what the composition of these nuclear RNAs were or how they connected to proteinaceous partners (Figure 1).

What are the hnRNAs that make-up this nucleoprotein rich structure? Some are highly abundant long **non-coding** RNAs (lncRNA) such as NEAT1 or MALAT1 or other diverse lncRNAs including FIRRE, HOTAIR, **XIST** etc. In contrast to these defined lncRNAs other studies [14,15] have suggested that a large proportion of nuclear RNAs hybridise to a **C_0t-1** probe, indicating that they are enriched in transcripts derived from a repetitive fraction of the genome. They also showed that these RNAs are quite stable and are turned over relatively slowly. However, it would be surprising if this was the complete repertoire of hnRNAs associated with chromatin when there are many other sources of RNA. Consistently, a rough analysis of chromatin associated RNAs suggested they derive from the intronic (52%), intergenic (19%) and exonic (29%) portions of the genome [16]. Interestingly similar values were obtained from estimates of RNA read abundance from RNA-seq data 36% exonic, 43% intronic, 20% intergenic [17]. Splicing is very tightly linked to transcription; analyses of long genes suggest that RNAs are spliced almost immediately after they are synthesised [18]. More recently RNA-seq data has been used to further quantify this using a completed splicing index (coSI) which decreases along genes, pointing to a “first transcribed, first spliced” rule [19]. By analysing the kinetics of splicing the Cook lab suggested that intronic RNAs were quite unstable, but they only looked at specific RNA species and it is possible that different RNAs are turned over at different rates [20]. In the cytoplasm mRNA half-life has a median time of 7.1 h [21] but in the nucleus RNA stability might be quite different. What might protect some RNAs from rapid degradation? Capped RNAs are more resistant to nuclease digestion so C_0t-1 RNAs might be protected compared to intronic RNA. Alternatively do RNAs adopt specific secondary structures that protect them from being degraded? **SHAPE** (selective 2’-hydroxyl acylation analyzed by primer extension) and computational approaches were used to map the structure of NEAT1 and showed it had extensive local base-pairing interactions and long range RNA-RNA interactions [22]. Local base pairing might provide protein binding sites to protect the RNA from degradation whilst the longer-range interactions are speculated to be important for paraspeckle formation [23].

RNA mobility

Macromolecules (e.g. globular proteins and small dextrans) are able to move freely through condensed chromatin domains [24], but the nucleus is a viscous environment, reportedly 300 times that of honey [25] so larger flexible molecules such as free RNAs are restricted. To put this into context a 1 kb RNA has a molecular weight of 330 kDa and a length of 340 nm. To facilitate transport, mRNAs have to be packaged into discrete ribonucleoprotein particles through a process that is highly coupled to transcription.

How do long non-coding RNAs move? In an elegant study to analyse the binding of the non-coding XIST RNA to chromatin the authors used RAP-seq (Box 2) [26]. After inducing the expression of XIST using doxycycline they showed that XIST RNA transfers to distal sites on the X-chromosome taking advantage of the 3D organisation and then propagates locally along the surface of the chromosome. Is XIST unique in this behaviour or could this be a general mechanism for lncRNA movement? NEAT1 is a highly abundant nuclear RNA required for forming paraspeckles perhaps through a **phase separation** process [27]. Does NEAT1 move in a similar way to XIST or is it folded into hnRNPs after synthesis and transported within the nucleus to the sites of paraspeckle formation? Another ncRNA located adjacent to NEAT1 is called MALAT1, both were observed to localise to hundreds of transcriptionally active chromatin sites [28]. Significant debate surrounds these abundant lncRNAs as NEAT1 and MALAT1 knockouts are viable, with no pronounced phenotype [29-31], although as the genes are conserved it would seem surprising if they didn't have some role that provides a selective advantage [32]. Molecular functions have been proposed for these genes including control of alternative splicing or transcriptional regulation. Another possibility is that high level transcription might alter the local chromatin structure or supercoiling and influence the surrounding transcriptional environment.

In contrast to defined lncRNAs, it has been shown that repetitive non-coding RNA species, labelled using C_0t-1 DNA as a probe, do not migrate far from their site of synthesis [14,33]. To us this suggests that many RNAs produced in the nucleus, unless they are proactively transported, will stay in the vicinity of the site of transcription, it was then suggested that once structural RNAs are embedded with chromatin/scaffold proteins they might become very stable [15]. Few studies have looked at the fate of spliced out introns, although the Lawrence group reported that spliced out introns drift into the nucleoplasm over time [14]. In contrast Wada et al [20] show there is degradation of RNA in the first intron while the polymerase is still transcribing the second, implying these are coupled processes and that different classes of RNA might behave differently.

Can nuclear RNA affect chromatin folding?

An important determinant of chromatin folding are the linker histones. In contrast to core histones they are mobile and bind the histone dyad through their globular domain [34] with their long C-terminal tail

draped along the chromatin fibre. Under physiological salt conditions linker histones promote folding of the chromatin into folded structures; in this state the compact chromatin sediments rapidly in a sucrose gradient, but if the histones are removed, unfolding the fibre, its sedimentation decreases [35]. Free RNA is highly negatively charged so it comes as no surprise that it can efficiently strip linker histones from chromatin. Purified chromatin readily binds to RNA [36], and chromatin structure is changed in the presence of RNA [37,38]. This raises an interesting question: what role does nuclear RNA have in regulating chromatin structure? Purified chromatin is often found bound to RNAs but at this level of chromatin organisation it is difficult to imagine how these nuclear RNAs could provide any specificity. Recent methods taking advantage of massively parallel sequencing have been developed to map RNA binding to chromatin (Box 3). In one of these studies ChaR-seq was developed to map RNAs bound to chromatin in *Drosophila*. Their analysis indicated that much of the genome is covered in small RNA species including snoRNAs and scaRNAs, but there seemed to be little specificity [39]. This might indicate that RNAs have a general role in regulating structure, or alternatively these short RNA species might just have a preference to associate with "sticky" positively charged histone proteins. This is not without precedence as shown by recent studies; proteins like PRC2 bind RNAs promiscuously [40,41].

RNA debris

Although many RNA species are transcribed from the genome by RNA polymerase II e.g. coding RNA, non-coding RNAs, snRNAs etc (Table 1), the species that is by far the most abundant is coding RNAs (mRNAs and pre-mRNAs) and can account for 75% of newly transcribed RNAs. As only approximately 5-7% of most coding genes are exons this follows that the intronic sequences would rapidly accumulate in the nucleus unless they are efficiently degraded. There are two main nuclear exonucleases XRN2 [42] and the exosome. Both of these are thought to have different roles and different specificities. XRN2 is a nuclease that recognizes single-stranded RNA with a 5'-terminal monophosphate and degrades it processively to mononucleotides, whilst the exosome complex degrades RNA starting at the 3' end. Structurally the exosome has been well characterised and looks like a molecular pencil sharpener with the RNA to be degraded threaded through the centre of the multi-subunit complex. XRN2 instead appears to work as a single subunit protein and is implicated in multiple roles within the nucleus including rRNA and snoRNA processing, transcriptional termination of specific genes and in human cells, XRN2 appears to be more important for degradation of aberrant pre-mRNA products than the exosome [43].

Long non-coding RNA half-lives vary widely and on average are only slightly less stable than mRNAs [44]. In contrast the half-life of intronic RNAs are turned over relatively rapidly with a median t1/2 <2.1 h [45,46] – but before degradation do they have a role in regulating chromatin structure? If they are able to interact with nuclear proteins, albeit transiently could they influence the organisation of a nuclear framework and

if they are rapidly turned-over could be considered as a special type of dynamic nuclear compartment. This is in contrast to how researchers often perceived the nuclear matrix - as a relatively static structure. Too few studies have yet investigated this in detail but the notion that nuclear RNA debris (e.g. intronic RNAs; Box 1) might play a role in regulating chromatin structure is appealing; it also fits well with the notion that many processes evolve to utilise by-products of other cellular events. Unfortunately, it is difficult to directly assess what role nuclear RNAs might have in regulating chromatin structure. One very blunt instrument is to prevent transcription by using small molecule inhibitors [47]. Transcription inhibition triggers chromatin compaction at the scale of 100's of Kb to the level of chromosomes territories [48,49]. Is this due to altered transcription *per se* or the loss of nuclear RNA? To dissect these possibilities, rather than perturb the level of nuclear RNA by either treating cells with exogenous nucleases or micro-injecting additional RNAs, it is easier and molecularly more controllable to ablate protein mediators of chromatin structure.

RNA/protein nuclear mesh

The fundamental nature and integral role of nuclear RNAs has made their study difficult. From early experiments which showed that a large component of chromatin is comprised of RNA to now, many of these studies have been largely indirect. In part this is because it is not possible to ablate nuclear RNA in a manner that is easy to report, and only very recently have imaginative new approaches been developed for quantitatively and qualitatively assessing the composition of nuclear RNAs (Table 1). Other than a very generic role in electrostatically competing off chromatin binding proteins such as linker histones we imagine nuclear RNAs will interact via protein mediators. This is borne out by formaldehyde RNA immunoprecipitation (fRIP-seq) that showed chromatin associated proteins bind thousands of RNAs [50]. Therefore, one route to understanding nuclear RNAs and the role they might play in regulating chromatin structure is by examining their partner proteins.

Scaffold attachment factor A (SAF-A) (Figure 2) was originally reported as an abundant protein component of a nuclear matrix [51] and was simultaneously found as a member of the family of hnRNPs [52]. It has a pronounced RG/RGG domain [53,54], an AAA+ domain for oligomerisation [33], and SAP and a SPRY domains of unknown function. SAF-A going by its other name HNRNP-U is suggested to have a role in regulating both XIST RNA binding to chromatin [55] and the splicing machinery [56]. HNRNP-U knockdown is reported to promote a misregulation of the SMN2 splicing factor which in turn had an impact on the splicing of a number of down-stream genes. Although a SAF-A/HNRNP-U hypomorph is early embryonic lethal [57], knocking HNRNP-U out in the heart gives rise to aberrant alternative splicing of a small number of genes [58]. However, in our hands depletion of SAF-A affects interphase chromatin structure but has little effect on transcription, as measured by RNA-seq and ribonucleotide incorporation [33]. Similar

results were observed by Fan et al [59] who showed that SAF-A has a global role in chromatin compaction and protein depletion caused significant alterations in TAD boundaries and reduced chromatin loop formation.

The abundant protein mediators that have been studied so far, such as SAF-A, do not appear to have sequence specific RNA binding motifs suggesting that interaction with nuclear RNAs is somewhat non-specific and quite week [53]. However, this is not to suggest this is an unregulated process. We have shown that SAF-A is able to oligomerise in the presence of RNA and biochemically seems to have an ability to form a nuclear mesh (Figure 3, Key Figure) [33]. Oligomerisation is ATP dependent through SAF-A's AAA+ domain and the protein has a number of putative phosphorylation sites that might also be regulatory. Furthermore, SAF-A has two protein paralogs (HNRNPUL1 and HNRNPUL2; Figure 2) that are structurally similar and might have a role in regulating oligomerisation or the length of oligomers, analogous to capping proteins for actin filaments. How SAF-A regulates chromatin structure is still unclear. Ablation of the protein or mutation of its Walker A motif in the ATP binding site inhibits protein oligomerisation and by microscopy compacts chromatin. There is a similar phenotype if the RNA binding domain is deleted suggesting that SAF-A oligomerisation is RNA-dependent and therefore responsive to local transcription, concomitantly the Rinn lab showed that SAF-A is preferentially bound to intronic RNAs [50]. Our experiments so far have suggested that the SAF-A/RNA mesh does not directly bind to chromatin but inhibits chromatin fibre-fibre interactions, modulating interphase chromatin structure. Hall et al likened the physical presence of the RNA to 'water' than can keep the chromatin/scaffold 'sponge' from shrinking (compacting) [15]. In *Drosophila* the abundant Df31 protein was identified as interacting with snoRNAs to regulate chromatin structure [38]. The mechanism is not clear but as it seems to function at transcriptionally active regions of the genome similar to SAF-A [33,59] it is tempting to speculate that the underlying mechanisms may be similar to how SAF-A is able to regulate structure. MATR3 is another abundant nuclear protein, like SAF-A it has been reported to bind introns through a pyrimidine-rich consensus sequence [60] and has a role in alternate splicing via the PTB regulator [61]. MATR3 is often referred to as a "nuclear matrix" protein, specific mutations give rise to amyotrophic lateral sclerosis through aberrant RNA processing [62] or neurotoxicity [63].

It is surprising that an alteration in large-scale chromatin structure, caused by SAF-A perturbation, does not appear to have a significant effect on gene transcription. This suggests that SAF-A's role in regulating chromatin structure is both down-stream of transcription and at a different level of genome organisation. Instead it might be that the process of local transcription and subsequent splicing generates a pool of intronic RNA debris that can transiently interact with SAF-A forming a nuclear mesh (Figure 5), regulating chromatin structure. This then implies that processes happening at a local chromatin scale such as

transcription could regulate large scale chromatin structures. It does however seem surprising that a local "nuclear mesh" wouldn't have some role in regulating transcription [64], not least because regions in the nucleus that have the highest levels of transcription, and concomitantly RNA debris, would be expected to form a more dense or extensive RNA-protein mesh. It is tempting to speculate that a local mesh or gel might provide a micro-environment for maintaining local transcription factor or regulatory protein concentrations for altering the efficiency of transcription [65-67].

If nuclear RNAs have a significant role in regulating chromatin organisation their function would be expected to be conserved going back through evolution. So far very few studies have investigated nuclear RNAs in species other than higher eukaryotes, however proteins like SAF-A which have been suggested to be structural mediators of genome organisation are conserved. By searching phyla for proteins with domains similar to SAF-A we identified orthologs in all Metazoa and SPRY – AAA domain structure is present in unicellular eukaryotes such as *Tetrahymena thermophila*. However, it was not possible to identify an ortholog in yeast so instead a SAF-A like function could be performed by two proteins that operate as a dimer. Alternatively, a rapidly growing transcriptionally active species might have less of a requirement for a global regulation of chromatin structure. Although SAF-A has not been well characterised evidence from databases indicate that both *C.elegans* and *Drosophila* mutants for the SAF-A are early embryonic lethal, suggesting they have an important role in cellular function. This is supported by new studies that report SAF-A mutations in patients with neurodevelopmental disorders [68-70]. The underlying molecular mechanisms are not understood but it is exciting to speculate that SAF-A haploinsufficiency might alter the structure or rigidity of a nuclear RNA mesh. Perhaps neural cells are particularly sensitive or instead altering protein/RNA interactions affects genomic stability [33], so during development some cells do not form correctly giving rise to a phenotype.

Concluding remarks

Despite many years of important experiments there are still many questions surrounding the role of a protein/RNA mesh (Figure 5) in regulating chromatin structure (see Outstanding Questions). It is clear that RNA is an abundant component of chromatin, and misregulation of RNA binding proteins such as SAF-A/hnRNP-U affect the structure of interphase and consequently mitotic chromosome structures. To better understand what is going on will necessitate more characterisation of the spectrum and turnover of RNA species in the nucleus and a better appreciation of the proteins that interact with these RNA species. Although biochemical properties of proteins like SAF-A/HNRNP-U are starting to be dissected we do not know what a nuclear mesh might resemble. Electron microscopy of extracted nuclei reveals a matrix like structure: what are its constituents? What does it look like in intact cells? New super-resolution microscopy techniques will facilitate these type of experiments and shed light on the dynamics of these

structures and new next generation sequencing techniques (Box 3) are being used to identify which RNA species interact with chromatin. These methods will need to be accompanied by new labelling techniques (Box 1) to enable nuclear RNA structures to be observed and visualised in response to different cell cycle stages or cellular perturbations such as transcription inhibition or the acute depletion of RNA binding proteins.

Acknowledgements

We thank members of our lab for many discussions about RNA binding proteins, nuclear mesh and chromatin architecture. R.-S.N. was supported by a Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science (JSPS) and N.G. is a UK Medical Research Council Senior Non-Clinical Fellow (MR/J00913X/1).

Figure 1. RNA/protein nuclear matrix. A cartoon representation of a nuclear mesh/matrix as visualised by electron microscopy following extraction of soluble proteins. A network of core filaments enmesh dense bodies (DB) that are rich in splicing components. Adapted from [71].

Figure 2. SAF-A/HNRNP-U domain composition and orthologs. Diagram showing SAF-A domains (SAP, SPRY, AAA+ and RG/RGG). HNRNPUL1 has an amino acid substitution in the Walker A motif within the AAA+ domain that is predicted to abolish ATP binding (SAF-A: GAGKT -> L1: AAGKT). HNRNPUL2 has very few RG/RGG motifs suggesting reduced RNA binding . The number of RG/RGG motifs for orthologs is SAF-A: 15, HNRNPUL1: 8, HNRNPUL2: 4.

F**igure 3. Key Figure. Formation of a protein/RNA rich mesh in the nuclei of mammalian cells.** SAF-A can oligomerise in the presence of RNA debris to form a nuclear mesh that maintains large-scale interphase chromatin in an open configuration.

Box 1

***In vivo* and *in vitro* RNA labeling methods**

Radioisotopes

For example [5-^{3}H] Uridine or [2-^{14}C] Uridine.

5BrU (5-Bromouridine)

5BrU is a uridine derivative with a bromo substituent at the fifth carbon. 5BrU-seq is based on BrU pulse labelling of nascent RNAs [72]. After the treatment with BrU in cells, BrU incorporated RNAs are immunoprecipitated with an anti-BrdU/BrU antibody in a pool of extracted total RNAs. However, it is reported that brominated pre-mRNAs are not good substrates for the cellular splicing machinery [73,74].

4sU (4-Thio uridine)

4sU is a photoreactive uridine derivative with a thiol substituent at the fourth carbon.
TT-seq (transient transcriptome sequencing) [75] uses 4SU to label nascent RNAs and enables estimates for the rates of RNA synthesis and degradation to be determined. 4sU incorporated RNAs are biotinylated with a thiol-reactive reagent and are purified following RNA fragmentation. The 4SUDRB-seq method [76] combines newly transcribed RNA labeling using 4sU with reversible inhibition of transcription elongation using 5,6-dichlorobenzimidazole 1-β-d-ribofuranoside (DRB). Immediately after release from DRB treatment, newly transcribed RNAs are tagged with 4sU and are converted to a sequencing library to measure genomewide transcriptional elongation rates and initiation frequency.

5EU (5-Ethynyl uridine)

5EU is a uridine derivative with an alkyne substituent at the fifth carbon [77]. 5EU-labeled RNA can be detected with a click reaction using a fluorescent azide.

Modified NTPs

GRO-seq (global run-on sequencing) uses 5Br-UTP to map the genome-wide distribution of transcriptionally engaged RNA pol II [78]. Isolated nuclei are incubated in a reaction including 5Br-UTP and newly synthesised RNAs containing 5Br-U are immune-precipitated with an anti BrU/BrdU antibody. Similar to GRO-seq, PRO-seq (precision nuclear run-on and sequencing) uses biotinylated ribonucleotide triphosphate analogs (biotin-NTP) instead of 5Br-UTP, which allows the efficient capture of newly transcribed RNAs for sequencing from their 3’ ends at base pair resolution.

Box 2

Techniques for mapping RNA bound to chromatin

MARGI (Mapping RNA genome interactions)

This technique has been used for mapping global RNA-chromatin interactions in human embryonic stem cells and human embryonic kidney cells. After extraction of formaldehyde/DSG cross-linked chromatin, RNA-DNA is bridged with a biotinylated double-strand oligo by proximity ligation, is purified and converted to a sequencing library for paired-end sequencing.

GRID-seq (Global RNA interactions with DNA by deep sequencing)

Similar to the MARGI method, GRID-seq method applies proximity ligation RNA to DNA with a biotinylated linker *in situ* on fixed chromatin, which reduces nonspecific interactions, in human, mouse and *Drosophila* cells. Biotin purified products are cleaned up with native polyacrylamide gel electrophoresis and are analysed using single-end 100 bp sequencing.

ChAR-seq (Chromatin-Associated RNA sequencing)

Similar to the MARGI and GRID-seq methods, ChAR-seq uses proximity ligation with a biotinylated oligo in situ on fixed chromatin in Drosophila cells [39]. ChAR-seq prepares relatively longer RNA and DNA fragments (20-100 bp) compared to the methods above and uses 152 bp long single-end reads to sequence across the entire junction of the bridge.

cheRNAs (chromatin-enriched RNAs). Stringent nuclear fractionation coupled to RNA sequencing. Purified nuclei from HEK293 cells are extracted with a 0.5M urea/ 0.5% NP-40 buffer to yield a soluble nuclear extract and insoluble chromatin pellet, and both pools are sequenced. Tightly chromatin associated lncRNAs identified from insoluble fraction are termed cheRNA.

DRIP-seq (DNA:RNA Immuno Precipitation)

DRIP-seq has been used for mapping DNA:RNA hybrid across the genome in human pluripotent Ntera2 cells using the monoclonal S9.6 antibody recognising DNA:RNA hybrids in a sequence independent manner.

Box 3

Techniques for analysing RNA-protein binding

CHART (Capture hybridization analysis of RNA targets)

CHART has been used to identify RNA bound DNA or protein partners in Drosophila cells [79]. This technique is a hybridization-based strategy that specifically enriches endogenous RNAs along with their targets with complementary oligonucleotides from reversibly cross-linked chromatin extracts.

ChIRP (Chromatin Isolation by RNA Purification)

Similar to the CHART assay, the ChIRP method is based on affinity capture of target lncRNA-associated DNA or proteins by biotinylated and tiling antisense-oligos [79]. The probe design requires prescreening and validation for maximum hybridization efficiency.

RAP (RNA Antisense Purification)

Similar to the CHART and ChIRP methods, RAP uses biotinylated antisense-oligos and was first used for determining the localization of Xist [26,79]. RAP uses 120-nucleotide antisense RNA probes to form extremely strong hybrids with the target RNA.

CLIP (Cross-linking immuneprecipitation), and its derivatives (HTS-CLIP, PAR-CLIP, iCLIP etc)

CLIP and its derivatives have been used to map the RNA sequences bound to an RNA binding protein in vivo with high resolution and specificity. Essentially, after UV crosslinking between RNA and its binding protein, the RNA–protein complex is immunoprecipitated using an antibody. The bound RNAs are ligated to an RNA linker, purified and analysed by sequencing.

Table1 Estimates of Pol II transcribed RNAs in mammalian cells, modified from [80]**.**

RNA Type	% total RNA by mass	Molecules per cell	Description
mRNA / pre-mRNA	66 – 99	5-20 × 10^5	Exonic, intronic and intergenic RNA
LINE1	0.022 – 0.066	3-10 × 10^2	Long Interspersed nuclear element-1. LINE-1 is classified in the non-LTR group of transposable element which is widespread in the genome of many eukaryotes. L1 elements have coding capacity. L1 is therefore classified as an autonomous retrotransposon.
SINE	0.066 – 0.22	9-30 × 10^2	Short interspersed elements. SINEs are also classified in the non-LTR group. SINEs do not encode for their own proteins necessary for transposition but can be mobilized by LINE-1-encoded proteins. SINEs that can be transactivated by LINE-1 are Alu and SVA elements.
LTR	0.033 – 0.11	4.5-15 × 10^2	Long terminal repeat. The LTR retroelement group includes endogenous retroviruses, whose genome organization parallels that of retroviruses.
Circular RNA	0.044 – 0.66	9-30 × 10^2	circRNAs are important players in normal cellular differentiation and tissue homeostasis as well as in disease development
snRNA	0.44 – 6.6	1-5 × 10^5	snRNAs play important roles in splicing of introns from primary genomic transcripts
snoRNA	0.88 – 4.4	2-3 × 10^5	snoRNAs form a large and abundant family of small noncoding RNA playing a conserved role in ribosome biogenesis
miRNA	0.066 – 0.44	1-3 × 10^5	miRNA is a subset of non-coding RNAs, which are ~22-nt long short RNA molecules that are considered to post-transcriptionally regulate the cleavage of target mRNAs or just repress their translation
XIST RNA	0.0066 – 4.4	0.1-2 × 10^3	Described in glossary
Other lncRNA	0.66 – 4.4	3-50 × 10^3	Additional defined non-coding RNAs

Glossary

C_0t-1: C_0t-1 DNA is enriched in repetitive DNA sequences such as LINEs and SINEs and is commonly used to block nonspecific hybridization. "C_0t" is experimentally defined as the product of DNA concentration at time zero (C_0) and time of reannealing and is used for C_0t analysis to fractionate populations of differently repetitive DNA.

FRAP: Fluorescence recovery after photobleaching. A method for measuring and analysing the recovery of fluorescence intensity in the microspace caused by the fluorescent molecules flowing into a region after photobleaching of fluorescence. The technique is often used for analysing protein mobility in cells.

hnRNA: heterogeneous nuclear RNA. hnRNA is often used as a synonym for pre-mRNA, however hnRNA also includes nuclear-retained RNAs that do not end up as cytoplasmic mRNA.

ncRNA: non-coding RNA. ncRNA is a generic term for RNAs that are not translated into proteins. They are further divided into small and long classes: small ncRNAs (sncRNA) being less than *200 nucleotides* and *lncRNA* (long non-codingRNAs) being greater than *200 nucleotides* to over 100 kb in length.

Nuclear matrix: The nuclear matrix model is proposed to organise interphase chromatin architecture. Nuclear matrixes are operationally defined ribonucleoproteinaceous structures that are resistant to high-salt buffers, nonionic detergents and RNA digestion. Experimental results for preparing the nuclear matrix are often not comparable.

Phase separation: A phase separation model proposes that multi-molecular assemblies such as RNAs and RNA binding proteins provide a general regulatory mechanism to compartmentalise chromatin or biochemical reactions within cells.

SHAPE: Selective 2'-hydroxyl acylation analyzed by primer extension. SHAPE is a method for probing RNA secondary structure. SHAPE reagents react preferentially with the 2'-hydroxyl groups of conformationally flexible RNA nucleotides. SHAPE reactivities are insensitive to base identity, and they correlate with local nucleotide flexibility and dynamics.

XIST: X-inactive specific transcript. XIST is a long non-coding RNA that remains associated with the X chromosome from which it is expressed. During initiation of X-chromosome inactivation XIST plays an essential role for recruiting a diverse set of proteins and epigenetic marks to establish a silent inactive X-chromosome.

References

1 Pederson, T. (2000) Half a century of “the nuclear matrix.” *Mol. Biol. Cell* 11, 799–805

2 Melé, M. and Rinn, J.L. (2016) ``Cat's Cradling'' the 3D Genome by the Act of LncRNA Transcription. *Mol. Cell* 62, 657–664

3 Caudron-Herger, M. and Rippe, K. (2012) Nuclear architecture by RNA. *Curr Opin Genet Dev* 22, 179–187

4 Jackson, D.A. *et al.* (1981) RNA is synthesized at the nuclear cage. *Nature* 292, 552–555

5 Jackson, D.A. and Cook, P.R. (1985) Transcription occurs at a nucleoskeleton. *EMBO J.* 4, 919–925

6 Jackson, D.A. *et al.* (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12, 1059–1065

7 Cisse, I.I. *et al.* (2013) Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* 341, 664–667

8 Caudron-Herger, M. *et al.* (2015) Dissecting the nascent human transcriptome by analysing the RNA content of transcription factories. *Nucleic Acids Res.* 43, e95–e95

9 Jacobson, A. and Peltz, S.W. (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* 65, 693–739

10 Jackson, D.A. *et al.* (2000) The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* 14, 242–254

11 Holmes, D.S. *et al.* (1972) Chromosomal RNA: its properties. *Science* 177, 72–74

12 Fey, E.G. *et al.* (1986) Association of RNA with the cytoskeleton and the nuclear matrix. *J. Cell Sci. Suppl.* 5, 99–119

13 Fey, E.G. *et al.* (1986) The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. *The Journal of Cell Biology* 102, 1654–1665

14 Hall, L.L. *et al.* (2014) Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* 156, 907–919

15 Hall, L.L. and Lawrence, J.B. (2016) RNA as a fundamental component of interphase chromosomes: could repeats prove key? *Curr Opin Genet Dev* 37, 137–147

16 Mondal, T. *et al.* (2010) Characterization of the RNA content of chromatin. *Genome Res.* 20, 899–907

17 St Laurent, G. *et al.* (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 13, 504

18 Singh, J. and Padgett, R.A. (2009) Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133

19 Tilgner, H. *et al.* (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Research* 22, 1616–1625

20 Wada, Y. *et al.* (2009) A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18357–18361

21 Sharova, L.V. *et al.* (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* 16, 45–58

22 Lin, Y. *et al.* (2018) Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res.* 46, 3742–3752

23 Bond, C.S. and Fox, A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* 186, 637–644

24 Verschure, P.J. *et al.* (2003) Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO reports* 4, 861–866

25 Caragine, C.M. *et al.* (2018) Surface Fluctuations and Coalescence of Nucleolar Droplets in the Human Cell Nucleus. *Phys. Rev. Lett.* 121, 148101

26 Engreitz, J.M. *et al.* (2013) The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* 341, 1237973–1237973

27 Yamazaki, T. *et al.* (2018) Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell* 70, 1038–1053.e7
28 West, J.A. *et al.* (2014) The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites. *Mol. Cell* 55, 791–802
29 Eißmann, M. *et al.* (2012) Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biology* 9, 1076–1087
30 Nakagawa, S. *et al.* (2012) Malat1 is not an essential component of nuclear speckles in mice. *RNA* 18, 1487–1499
31 Nakagawa, S. *et al.* (2011) Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* 193, 31–39
32 Nakagawa, S. *et al.* (2014) The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* 141, 4618–4627
33 Nozawa, R.-S. *et al.* (2017) SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin- Associated RNAs. *Cell* 169, 1214–1227.e18
34 Garcia-Saez, I. *et al.* (2018) Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. *Mol. Cell* DOI: 10.1016/j.molcel.2018.09.027
35 Allan, J. *et al.* (1981) Regulation of the higher-order structure of chromatin by histones-H1 and histones-H5. *The Journal of Cell Biology* 90, 279–288
36 Rodríguez-Campos, A. and Azorín, F. (2007) RNA is an integral component of chromatin that contributes to its structural organization. *PLoS ONE* 2, e1182
37 Schubert, T. and Längst, G. (2013) Changes in higher order structures of chromatin by RNP complexes. *RNA Biology* 10, 175–179
38 Schubert, T. *et al.* (2012) Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol. Cell* 48, 434–444
39 Bell, J.C. *et al.* (2018) Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* 7, 429
40 Davidovich, C. *et al.* (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* DOI: 10.1038/nsmb.2679
41 Davidovich, C. *et al.* (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* 57, 552–558
42 Miki, T.S. and Großhans, H. (2013) The multifunctional RNase XRN2. *Biochem. Soc. Trans.* 41, 825–830
43 Davidson, L. *et al.* (2012) Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J.* 31, 2566–2578
44 Clark, M.B. *et al.* (2012) Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22, 885–898
45 Ayupe, A.C. *et al.* (2015) Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biology* 12, 877–892
46 Hesselberth, J.R. (2013) Lives that introns lead after splicing. *Wiley Interdiscip Rev RNA* 4, 677–691
47 Bensaude, O. (2014) Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? *Transcription* 2, 103–108
48 Naughton, C. *et al.* (2010) Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol. Cell* 40, 397–409
49 Naughton, C. *et al.* (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* 20, 387–395
50 Hendrickson, D.G. *et al.* (2016) Widespread RNA binding by chromatin-associated proteins. *Genome Biol* 17, 28
51 Romig, H. *et al.* (1992) Characterization of Saf-a, a Novel Nuclear-Dna Binding-Protein From Hela-Cells with High-Affinity for Nuclear Matrix Scaffold Attachment Dna Elements. *EMBO J.* 11, 3431–3440
52 Kiledjian, M. and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* 11, 2655–2664

53 Thandapani, P. *et al.* (2013) Defining the RGG/RG Motif. *Mol. Cell* 50, 613–623
54 Ozdilek, B.A. *et al.* (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* 45, 7984–7996
55 Hasegawa, Y. *et al.* (2010) The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* 19, 469–476
56 Xiao, R. *et al.* (2012) Nuclear Matrix Factor hnRNP U/SAF-A Exerts a Global Control of Alternative Splicing by Regulating U2 snRNP Maturation. *Mol. Cell* 45, 656–668
57 Roshon, M.J. and Ruley, H.E. (2005) Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res.* 14, 179–192
58 Ye, J. *et al.* (2015) hnRNP U protein is required for normal pre-mRNA splicing and postnatal heart development and function. *Proc. Natl. Acad. Sci. U.S.A.* DOI: 10.1073/pnas.1508461112
59 Fan, H. *et al.* (2018) The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes. *Genome Res.* 28, 192–202
60 Uemura, Y. *et al.* (2017) Matrin3 binds directly to intronic pyrimidine-rich sequences and controls alternative splicing. *Genes Cells* 22, 785–798
61 Coelho, M.B. *et al.* (2015) Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *EMBO J.* DOI: 10.15252/embj.201489852
62 Johnson, J.O. *et al.* (2014) Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat. Neurosci.* 17, 664–666
63 Malik, A.M. *et al.* (2018) Matrin 3-dependent neurotoxicity is modified by nucleic acid binding and nucleocytoplasmic localization. *Elife* 7, 602
64 Hnisz, D. *et al.* (2017) A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23
65 Sabari, B.R. *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, eaar3958
66 Cho, W.-K. *et al.* (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415
67 Chong, S. *et al.* (2018) Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361, eaar2555
68 Yates, T.M. *et al.* (2017) De novo mutations in HNRNPU result in a neurodevelopmental syndrome. *Am. J. Med. Genet. A* 173, 3003–3012
69 Leduc, M.S. *et al.* (2017) Clinical and molecular characterization of de novo loss of function variants in HNRNPU. *Am. J. Med. Genet. A* 173, 2680–2689
70 Bramswig, N.C. *et al.* (2017) Heterozygous HNRNPU variants cause early onset epilepsy and severe intellectual disability. *Hum. Genet.* 136, 821–834
71 Nickerson, J.A. *et al.* (1997) The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4446–4450
72 Paulsen, M.T. *et al.* (2013) Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2240–2245
73 Sierakowska, H. *et al.* (1989) Inhibition of pre-mRNA splicing by 5-fluoro-, 5-chloro-, and 5-bromouridine. *J. Biol. Chem.* 264, 19185–19191
74 Wansink, D.G. *et al.* (1994) In vitro splicing of pre-mRNA containing bromouridine. *Mol. Biol. Rep.* 19, 109–113
75 Schwalb, B. *et al.* (2016) TT-seq maps the human transient transcriptome. *Science* 352, 1225–1228
76 Fuchs, G. *et al.* (2014) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* 15, R69
77 Jao, C.Y. and Salic, A. (2008) Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15779–15784
78 Core, L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848
79 Simon, M.D. (2016) Insight into lncRNA biology using hybridization capture analyses. *Biochim. Biophys. Acta* 1859, 121–127

80 Palazzo, A.F. and Lee, E.S. (2015) Non-coding RNA: what is functional and what is junk? *Front Genet* 6, 2

Figure 1

Click here to access/download;Figure;Fig1.tif

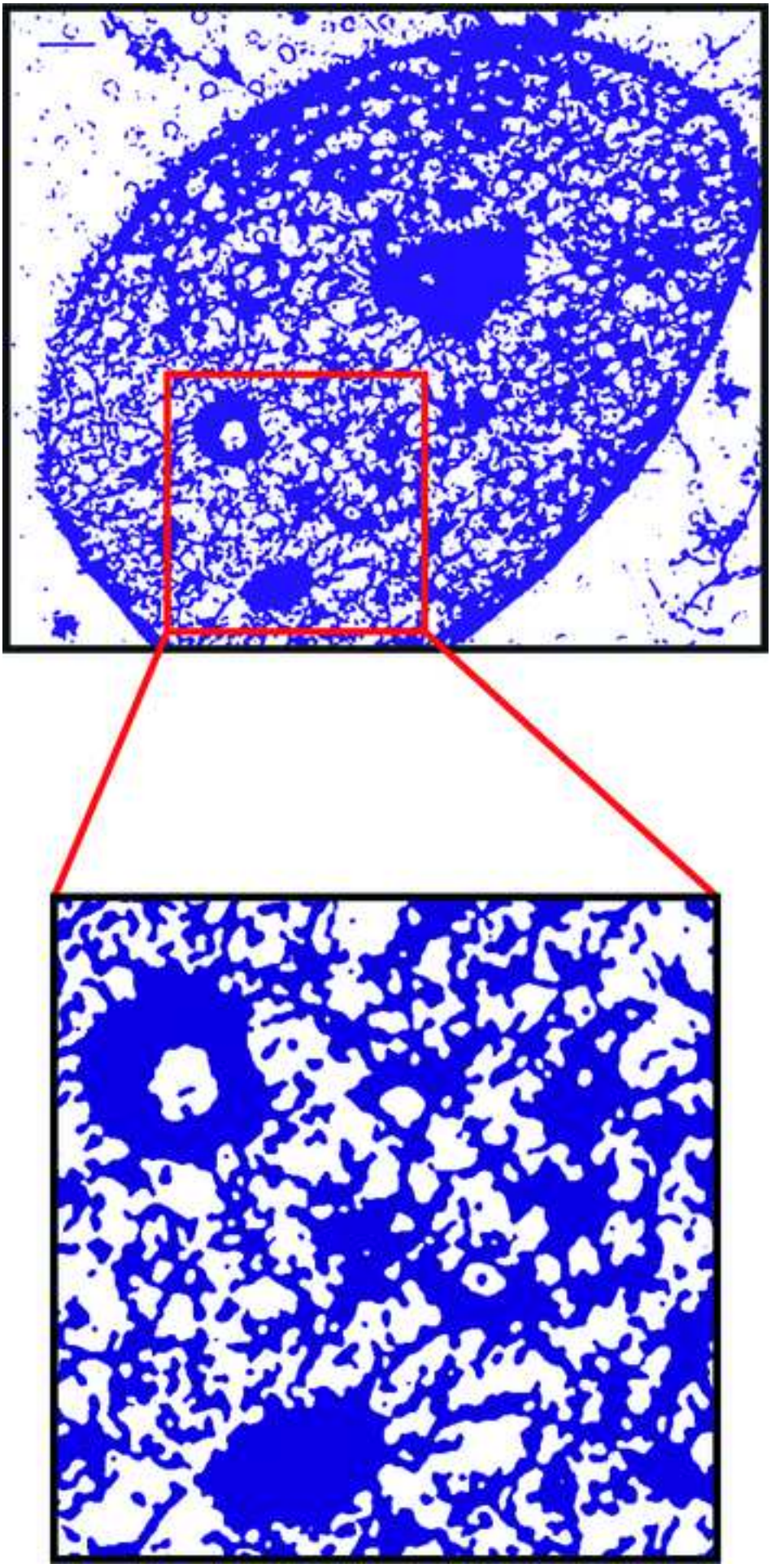
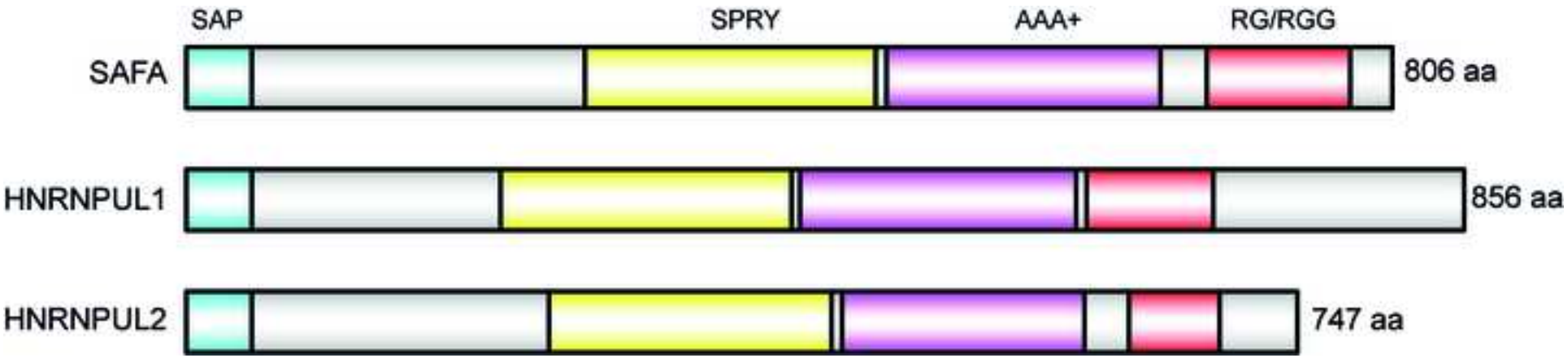
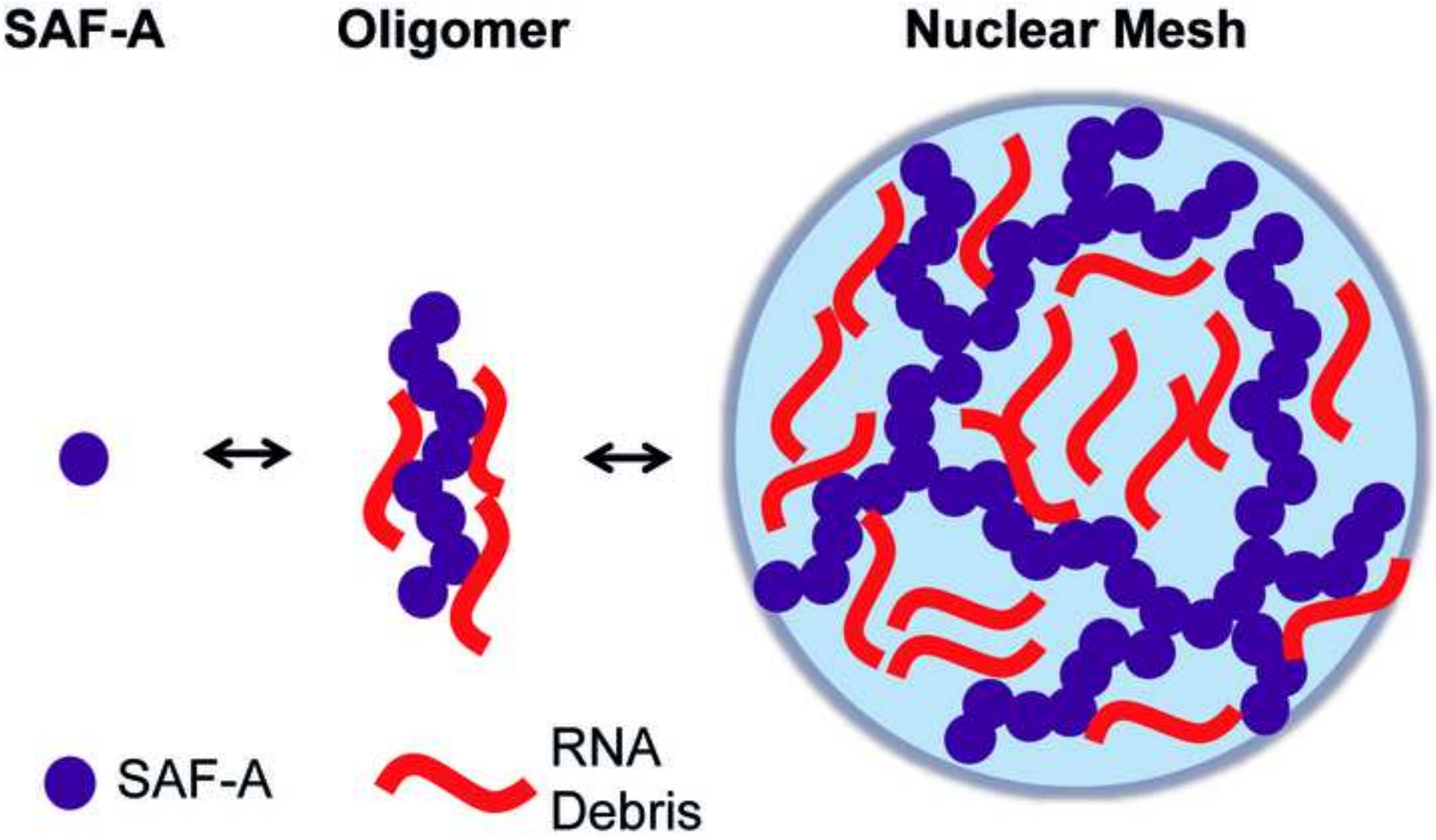


Figure 2





Outstanding questions

- The old idea of a static nuclear matrix is not compatible with our better understanding of how quickly processes occur in cells. How dynamic is a nuclear mesh? Does in vary between cells? Can we use the new approaches to address these important questions?
- We are starting to understand the biochemistry of a nuclear mesh but methods are needed for visualising an RNA/protein mesh in cells. New super-resolution imaging techniques will enable diffusion unlimited analysis of these complex structure and advances in cryo-EM tomography might complement these approaches.
- To visualise the components of a nuclear mesh we need new reagents. Antibodies for detecting specific protein modifications and new labelling techniques for marking hnRNAs for microscopy or biochemistry.
- What are the factors that regulate nuclear RNA binding proteins? Are these proteins regulated by phosphorylation?
- SAF-A is an AAA-domain containing protein. Does the protein oligomerise to form a spiral? Can we use knowledge from the crystallisation of other AAA-domain containing proteins to do X-ray crystallography? Will this help us to understand how these proteins work?
- Do SAF-A paralogs (e.g. HNRNPUL1 and HNRNPUL2) regulate protein oligomerisation? How does RNA interact with SAF-A? To gain new insight into how a nuclear mesh can regulate chromatin structure can we reconstitute it *in vitro* and image using single particle cryo-EM?
- Recent inspection of the cancer genome atlas indicates that sites within SAF-A are mutation hotspots in cancer. How does altering SAF-A function affect disease progression or tumour survival?
- Advances in CRISPR/Cas9 will enable better cell biology approaches for triggering the acute degradation of proteins important for nuclear mesh formation, or the introduction of patient specific mutations.

Highlights

- Nucleus is rich in hnRNA species, including defined long non-coding RNAs and various other RNA “debris” (e.g. introns, C_0t-1 RNA)
- Only 5% of polymerase II synthesised RNA is exported to the cytoplasm, presumably the remainder plays a role or is degraded in the nucleus
- Abundant nuclear proteins such as SAF-A/HNRNP-U non-specifically interact with RNA “debris”
- protein/RNA structures form a dynamic nuclear mesh, its role is not fully understood but it can regulate interphase chromatin structure